Location-based Trust for Mobile User-generated Content: Applications, Challenges and Implementations

Vincent Lenders, Emmanouil Koukoumidis, Pei Zhang and Margaret Martonosi
Dept. of Electrical Engineering
Princeton University
{lenders, ekoukoum, peizhang, mrm}@princeton.edu

ABSTRACT

The recent explosion in shared media content and sensed data produced by mobile end-users is challenging well-established principles and assumptions in data trust models. A fundamental issue we address in this paper is how to establish some trust level in the authenticity of content created by untrusted mobile users. We advocate a secure localization and certification service that allows content producers to tag their content with with a spatial timestamp indicating its physical location. At the same time, however, our approach preserves the privacy of producers by not exposing their identity to the potential content consumers. We provide a list of existing and possible applications that would profit from such a secure localization service and sketch possible implementations of the service, highlighting advantages and drawbacks.

1. INTRODUCTION

With recent advances in Web 2.0 technologies and the insatiable desire of people to share information, we are currently observing a paradigm shift in the way content is created and consumed. Whereas news items, photographs and other data items have traditionally been provided by a small group of professionals and consumed by a large audience, technology today allows more and more content to be provided by the mobile users themselves, for a broad community of people with common interests. This *user-generated content* is provided in the form of podcasts, blogs, or collaborative platforms such as Flickr [14], YouTube [26] or Wikipedia [21]. This new style of content sharing enables previously unimagined opportunities, but also requires some re-thinking of well-established principles.

A major concern that has been raised for user-generated content is how to trust the authenticity and quality of information that has been published by individuals, possibly mobile and often unknown to the content consumers. Traditionally, a news article published, for example, on a well-known website like Reuters or the BBC carried some implicit quality and authenticity guarantees based on the reputation of the news provider. In contrast, content from citizen journalists may come from multiple unknown individuals and may be published on possibly untrusted sites; these do not naturally embody such a priori confidence in the contents.

To provide a consumer with a level of trust for content provided by mobile users (e.g., a picture, a video, an audio, text, a sensor reading, etc.) from an unknown source, we would like a basic verification primitive to check the validity of contents. Ideally, such a primitive should not be proprietary to one specific application or type of content, but rather provided as a basic service to support a wide range of contents and applications to potential users.

The approach we advocate in this work is to couple the content with a spatial timestamp noting a system-verified time and location. This is similar to *geotagging*, consisting of adding geographical metadata to media. For example, photo sharing websites like Flickr allow geotagging to annotate the locations of pictures taken from different people. In this paper, we propose a trusted geotagging service: it should be very difficult for users to apply arbitrary tags for locations they have never been to. Trusted geotagging adds a clear benefit to the consumers of mobile user-generated content since they can now have greater confidence as to where some content has been, without having to personally know who created the content. Tagging data with location information that is difficult to fake is not directly possible when using traditional identity-based security, in which the identity of the content creator is associated with the content. Furthermore, identity-based security implies a trust relationship between the content providers and consumers; this is hard to achieve, in a large open system. In addition, the system would be prone to Sybil attacks [12] in which a malicious user propagates false content using multiple identities. These kinds of attacks are, however, greatly reduced in our location-based model since a user would need to physically move to a specific location in order to tag the content with that desired location.

We start by describing potential applications that would profit from location-based security and a secure geotagging service. Then, we present in Section 3 our system design. In Section 4, we describe possible implementations and conclude in Section 5.

2. APPLICATIONS

User generated content is more valuable (or in some cases, only useful) when its spatial and temporal properties can be verified. Such properties primarily involve the origin location and time of creation of the content. In other cases, it is also useful to know the path some content has traveled in order to reach the consumer. In this section, we review different types of applications that would profit from location and time verification of content.

HotMobile'08, February 26-26, 2008, Napa Valley, CA, USA
Copyright 2008 ACM 978-1-60558-118-7/08/0002 … $5.00.

News: News articles, videos or podcasts are typically describing the happenings at a particular location and time. The authenticity of the news report (i.e., pictures or description of a scene) could be verified by checking that the creation of the news report and media contents correlate to the location and time of the events.

Photo sharing: Photo sharing websites allow communities of people to share personal photographs. Flickr, one of the most prominent examples, is widely used by bloggers to share geo-tagged photographs. This allows one to search for photos that were taken in a particular geographical area. By certifying the location at the moment of the photo's creation, such sites would be able to automatically filter authentic pictures.

Distributed sensing: Applications are emerging that draw on sensed information about people, animals, objects, or physical spaces [2]. The data readings in such sensor networks are more valuable and trustworthy if they can be related to where and when the readings originated. The emerging network of amateur weather stations [20] is an example of this. By certifying the locations of data, it is much harder for potential attackers to fake readings.

Filtering email spam: Spam emails are typically spoofed by using a variety of different source email addresses. Their growing sophistication makes it a challenge for current spam filters to efficiently distinguish spam from real messages. A possible filtering rule would therefore consist of flagging emails that originate from the location of known spamming networks as well as emails with false location claims. For spammers that use the resources of a botnet, they would need to additionally match the location of machines they take over, with source email address they are spoofing.

Mapping data: Over the last couple of years there has been a huge move towards the inclusion of user-generated content in mapping data. Google, Tele Atlas (Tomtom) and Everyscape are using legions of GPS-empowered photographers to acquire the mapping data they need. In such cases location verification could prove very useful for trusting the contributions that users are making.

Traffic updates: Location plays a significant role in vehicular applications like traffic congestion detection and warning systems. In these systems, it is critical to determine that a message did indeed originate from a given location. For example, this primitive would prevent an attacker from injecting false traffic updates while not physically present at that location.

3. VERIFICATION SYSTEM FOR MOBILE CONTENT

3.1 Design goals

Our envisioned location-based certification system for user-generated content has the following design goals:

- Generality: We would like the service to be generic in order to support many types of contents and applications.

- Scalability: The service should scale to a large number of users as we envision that any end-user could be a potential content producer or consumer.

- Retains user privacy / anonymity: Users (content producers and consumers) should preserve their privacy and remain anonymous if desired. This is particularly critical if the service becomes ubiquitous and used by a majority of the population. It should not be possible to track individual users by looking at the content location tags

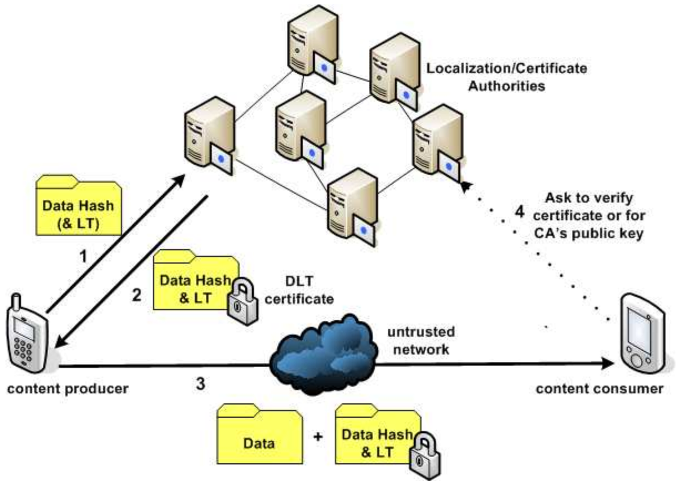

Figure 1: Content verification system overview.

- Support for user mobility: In some cases, information like pictures or videos will be generated and possibly consumed while people are mobile. The service should hence support mobility of the users.

- Support for content mobility: The content may be distributed over many different networks and channels. We want to still be able to verify the data even after it has been moved to another network location or site.

3.2 System Overview: Tracking Content not People

Our proposed system consists of three entities: content producers, content consumers, and a location/time verification service. The content producers and consumers are typically mobile users with a wireless device like a PDA, mobile phone, or a laptop. When a content producer (like the smartphone in Figure 1) has some content that it wants to have geographically certified, it issues a request to the localization/certificate authority (step 1). This request includes a hash over the content it wants to have certified. There are no restrictions on the particular hash function and hash index size used, as long as the function virtually guarantees that different contents will produce different hash indexes and that it is virtually impossible to create a desired content given the hash index. Example hash functions are the SHA hash function family [18]. Upon reception of the request, the localization/certificate authority performs two tasks. First, it determines the location of the mobile content producer in a secure way, i.e., in a way that the content producer cannot spoof its location (see Section 4 for technical implementations of this problem). Then, it replies back to the content producer with a Data-Location-Time (DLT) certificate (step 2) that binds the location of the content producer with the current time and the hash of the content. For privacy reasons, the content producer has the option to blur its location by specifying the level of accuracy for the location and time information in the certificate. For example, it could request to have city/day value rather than a precise coordinate and time value with second precision. The level of precision is entirely controlled by the content producer. An important aspect is that the certificate does not include the identity of the content producer. It is hence not possible to recover the identity of the content producer from the DLT certificate.

The content producer now has the option to publish its content with the issued certificate (step 3). The content will typically be copied to a server in the Internet or directly sent to the content producer by any traditional delivery mechanism like for example emails or possibly by means of a wireless ad hoc network among the mobile users. When the content consumer retrieves the content, it can now verify the origin location and time of the content by verifying the authenticity of the certificate (step 4), i.e., by verifying the signature of the certificate using the public key of the localization/certificate authority. Our system is intended for use to improve confidence at the generation or source of the data. Therefore, it does not allow modification of this data after generation. We will discuss how to implement the localization/certificate authorities in the next section and continue with general systems aspects next.

3.3 Trust Model

Given the large number of possible content consumers and producers, we do not assume any trust relationship between the consumers and the producers in the system. The content producers do not have to trust the localization/certificate authorities for correctly determining their location since they can verify themselves the certificate after it is issued. In fact, producers could even propose a location and time value that authorities sign if correct. The content consumers, however, trust the authorities that they only issue certificates with verified locations of the content producer and correct time. In order for the content producers to keep their privacy and remain anonymous, they further trust the authorities that they do not store nor reveal any information about received certificate requests. Such information could potentially be used to track a user's mobility pattern.

3.4 Privacy Issues

As highlighted earlier, content producers are exposed to potential tracking attacks since they reveal their location in the certificates for their contents. This might be a privacy concern when a producer has lots of contents or when the location at which the content is created is not heavily populated. For this reason, the content consumers have the ability to specify the level of accuracy and anonymity for the location and time values in the certificates. They can specify the highest accuracy for applications in which privacy is not a concern and request blurred coordinates and time values in other cases.

3.5 Possible Attacks

In addition to possible low-level attacks on the underlying cryptographic mechanisms or attacks on the location verification of the wireless nodes, we review here three possible attacks to the system as a whole and how they can be mitigated.

Mobility attack: In this attack, a malicious user that wants to issue some content with false location information in the DLT certificate would move to that particular location and obtain the certificate. The system cannot directly prevent or detect such attacks. However, the cost for these types of attacks is extremely high since it requires physical movement of the attacker to that location. The attacker is limited by the time it takes to travel to a specific position.

Botnet attack: The Botnet attack consists of relaying a certificate request through compromised zombies that an attacker controls [8]. The zombies are selected by the attacker according to the geographical region of interest to obtain a false certificate. However, it is difficult for an attacker to identify compromised nodes required for each different geographical regions. Furthermore, with the geographical constraint limiting the number of zombies available to the attacker for a given attack, this system can more easily identify invalid high-confidence certificates from Botnets.

Delay attack: This attack consists of querying for the DLT certificate a certain amount of time after the creation of the content to make the impression that the content was created there. This type of attack can be detected if the receiver has some expectations about the creation time of the content, as in news. In that case, the time stamp in the certificate would reveal a possible abuse.

3.6 Mobility Considerations

A major concern with mobility is when the content producers are moving while they generate content and request for certificates. This type of mobility makes it challenging to assign a unique position to the content at high speeds. For example, the recording of a video might take too long relative to the user's mobility to be able to assign it with one unique position. In such cases, we propose to assign multiple DLTs to contents, allowing one to track the mobility of the content over its creation time. This scheme implies that a video is accompanied by a list of DLTs. Still, the size of the list will be small since certificates include a hashed index of the content that is relatively small in size. In addition, such a scheme could further be used in general to track the mobility of content even when the content creation time is small. It would then allow to a content consumers to know over which path some content has been moving before being delivered.

3.7 Alternative Certification Mechanisms

The certification mechanism described in Section 3.2 draws a lot from the way digital signatures work [22, 24, 25]. Our proposed verification service serves the role of the signer of the user-generated data after appending location and time information to it. In another sense it is similar to the way trusted digital timestamping work according to the RFC 3161 [5] or the newer ANSI ASC X9.95 standard [3]. But again in our case the timestamping authority needs to have also secure localization capabilities and include also location information in the stamps.

An alternative way of certifying user-generated content is to use DLT servers that store the hash of the data and also their corresponding location and time information. In such a scheme, every time a user wants to make its content certifiable, it issues a request to the DLT server by sending the hash of the data. The server on receiving the hash will securely determine the location of the device and also record the time. The consumer of the content can later contact the server to verify the location and time information of a piece of data by just providing its hash. This idea draws from the way Bit Torrent [4] verifies the integrity of the data that users are downloading.

Our system does not support editing of certified data. This is by design, because in many cases, such edits are undesirable and our authentication system aims to increase the confidence that the data is unaltered. However, in some cases, editing content (such as cropping a picture) is desired. Our system may be extended to handle this. For example, one might incorporate changes with a difference list, where the users decide if the changes are acceptable. This scheme again necessitates the existence of a verification service that is calculating and embedding the watermarks, allowing for the ability to verify the certificate. . .

4. LOCALIZATION/CERTIFICATE SERVICE: IMPLEMENTATION

Our system design requires a secure localization and certificate issuing service. This section describes several challenges that arise with the implementation of such a service.

4.1 Secure Wireless Localization

There are many wireless localization techniques that can be used to obtain the location of a mobile device. However, the most popular techniques allow one to determine one's own position and are often not secured against spoofing attacks. We next review research work that aims at alleviating these two limitations.

Satellite-based positioning (e.g., GPS) offers meter accuracy positioning almost everywhere on the planet. However, this type of positioning is passive, meaning that the mobile devices determine their own position and the satellites cannot determine the location of mobile devices on earth. A possibility would be to have trusted nodes (fixed or mobile) which are equipped with GPS and are operated by a trusted party to perform the localization with the content producers using a small range wireless technology. An alternative interesting approach has been proposed in [10]. With this method of secure localization using GPS, the physical location of a particular node can be verified by a close-by third node on earth. The verification is based on the signature of the received signals from the different satellites which offers enough randomness to make it virtually impossible to forge. This would allow neighboring nodes to certify a node's claimed GPS position without the need for other wireless short-range localization techniques.

Another form of localization is tower localization. This method uses multiple cellular phone towers to centrally locate a node's position. In regions with a high tower density as in highly populated urban areas, such a scheme can provide up to meter precision even indoors [17]. This method could be offered as a service by the existing telephone providers. Also given the high density of 802.11 access points, one could use those to determine the location of content consumers in a similar way [16]. Since such localization systems are prone to location spoofing attacks, they must be secured using with additional techniques (e.g., [6]).

A hybrid version of these two aforementioned localization schemes is performed by Yahoo!'s ZoneTag cellular application for tagging pictures [1]. ZoneTag first tries to acquire its location information by a GPS device that may be attached to the cell phone. If no such device exists, it tries to infer its location by information that other users have inserted about the location of the cell that it currently lies in. This means that other users must have used in the past the ZoneTag application within the same cell and provided valid location information. However, this just helps localization for the purpose of tagging pictures and does not increase trust especially since users can also manually edit their location tag. Cellular telephone towers should be able to behave as certificate authorities or proxies for those in order to establish trust and have a more secure way of establishing their own location that can not be manipulated by the users.

There have been generally several other systems [7] [15] [19] designed for network-based geolocation. However, none of these has been targeted for secure tagging of user-generated content.

4.2 Certificate Authorities

In principle, the entities that issue DLT certificates could delegate the location verification tasks to other parties. However, for security and ease of deployment reasons we envision a system in which the authorities that perform the verification also issue the certificates. We next review different possible implementation of the certificate authority.

Centralized: The most obvious method for obtaining a certificate is through a centralized location server. In such a scenario, a central server is queried whenever the content generator requires a DLT certificate. The central server is responsible for verifying the location that the generator is claiming to have, and responding with the corresponding DLT certificate. In such a system, implicit trust is given to the central authority. In addition, a central server provides the simplest method for a limited system implementation. The central server method also has a few drawbacks. It implicitly assumes that the content producer has access to the server, for example through a backbone network or the Internet. When this is unavailable however, the device could receive a not-useful DLT certificate. For example, consider content recorded by a reporter in Burma with a device that is not connected until it reaches Germany. Furthermore, because our proposal attaches DLT certificates to content, the request could be frequent enough to allow the central authority to track a user and invade on one's privacy. In addition, a central server would also be much more susceptible to denial of service attacks.

Decentralized: One way to alleviate the problems associated with the centralized approach is to use an entirely distributed method. Such a method could include users asking multiple devices to obtain multiple DLT certificate, the generating device then will keep track of the trustworthy level of these other nodes. The receiver of the data can then query the DLT certificate source nodes to validate the data, or simply work offline if the certificate sources are previously known. In such a system, trust is placed in the public. An assumption is made that most devices are well behaved and privacy of users can be preserved by diluting the certificates obtaining process through multiple nodes, therefore no single node can keep track of any other node simply through their certificate requesting patterns. This method can be implemented in a similar way as Pretty Good Privacy (PGP [27]), one modification would be that the the public key is device specific instead of user name specific. The advantage of such a system is that it is more scalable than the centralized method. In addition any denial of service attacks would have very limited impact of such a system. Because of the multiple certificates obtained, it is difficult to fake an authentication. However the overhead of such a system would be significantly higher due to multiple parties that need to be contacted. This often provides multiple parties, especially trusted parties, to attack. The scopes of such attacks are more limited. Furthermore, when the data is important enough, the DLT certificate can still be faked with a community of fake devices as in Sybil attacks [12]. This vulnerability is difficult to circumvent especially if the community of fake devices become large. In addition, to verify the validity of a certificate from an unknown source, the receiver of the data needs to be online, and would not be able to connect when the device is offline. This requirement would limit the usefulness of the system if the device is not well connected.

Community Model: In the community model, a community of devices will vote and decide on the trust level of the certificates. Similar to the distributed method, this method uses multiple certificate issuers to allow for multiple certificates, however, these certificate givers are voted on by its users. These issuers will have different weight and trust levels associated with them. In addition, each user is also given votes to weigh its votes. This is similar to the community voting used by Digg [11] or reputation schemes used in eBay [13]. The community gets to vote on each other's trust level to issue the certificates and super users can boost or reduce this level.

5. CONCLUSIONS

It is beneficial for many applications in which mobile users create content to be able to verify the origin location and time of the content. We have proposed a secure localization and certification service in order for content consumers to establish the trust level of contents. The proposed system architecture maintains user privacy by tagging content with the location and time rather than the identity of the user that has generated the content. The system increases the difficulty for attackers to tag content with false location information as it requires them to physically move to that location. This limits the scalability and reduces the ease of many attacks and shows the usefulness of a location-based content verification service for user-generated data.

6. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful feedback on this work. Vincent Lenders was supported by a post-doctoral fellowship provided by the Swiss Government. In addition, we acknowledge the financial support of the National Science Foundation (Grant numbers 06-14949 and 02-05214).

7. REFERENCES

[1] S. Ahern, M. Davis, D. Eckles, S. King, M. Naaman, R. Nair, M. Spasojevic, and J. Yang. ZoneTag: Designing Context-Aware Mobile Media Capture to Increase Participation. In *Workshop on Pervasive Image Capture and Sharing (PICS 2006), Ubicomp 2006.*, 2006.

[2] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Survey on sensor networks. *Communications Magazine, IEEE*, 40:102–114, 2002.

[3] ANSI. Trusted Time Stamps, American National Standard X9.95–2005, 2005.

[4] Bit Torrent. http://www.bittorrent.com/.

[5] C. Adams, P. Cain, D. Pinkas and R. Zuccherato. Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP), The Internet Engineering Taskforce RFC3161–2001, 2001.

[6] S. Capkun and J. Hubaux. Secure positioning of wireless devices with application to sensor networks. In *Proceedings of IEEE INFOCOM*, Miama, FL, USA, March 2005.

[7] Cell-Loc. http://www.cell-loc.com/.

[8] E. Cooke, F. Jahanian, and D. Mcpherson. The Zombie Roundup: Understanding, Detecting, and Disrupting Botnets. In *Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI)*, pages 39–44, June 2006.

[9] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker. *Digital Watermarking and Steganography*, volume 1. Morgan Kaufmann, 2 edition, 2008.

[10] D. E. Denning and P. F. MacDoran. Location-Based Authentication: Grounding Cyberspace for Better Security. *In Elsevier Computer Fraud and Security*, February 1996.

[11] Digg. http://digg.com/.

[12] J. Douceur. The sybil attack. In *International Workshop on Peer-to-Peer Systems (IPTPS)*, Cambridge, MA, USA, 2002.

[13] eBay. http://www.ebay.com.

[14] Flickr. http://www.flickr.com/.

[15] Geometrix MLC. http://www.geometrix911.com/.

[16] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki. Practical robust localization over large-scale 802.11 wireless networks. In *Proceedings of ACM MOBICOM)*, Philadelphia, PA, USA, September 2004.

[17] V. Otason, A. Varshavsky, A. LaMarca, and E. de Lara. Accurate GSM Indoor Localization. In *Proceedings of UbiComp*, pages 141–158, Tokio, Japan, September 2005.

[18] B. Schneier. *Applied Cryptography: Protocols, Algorithms, and Source Code in C*. Wiley, 1996.

[19] TruePosition. What is trueposition cellular location system?, http://www.trueposition.com/. http://www.trueposition.com/intro.htm, 2001.

[20] Weather Underground. http://www.wunderground.com/.

[21] Wikipedia. http://wikipedia.org/.

[22] M. Wu and B. Liu. Watermarking for image authentication. In *In the proceedings of IEEE International Conference on Image Processing (ICIP'98)*, 1998.

[23] M. Wu and B. Liu. Watermarking for image authentication. In *International Conference on Image Processing (ICIP 98)*, Chicago, USA, October 1998.

[24] M. Wu and B. Liu. Data Hiding in Image and Video: Part-I – Fundamental Issues and Solutions. *IEEE Trans. on Image Proc.*, 12:685–695, 2003.

[25] M. Wu, H. Yu, and B. Liu. Data Hiding in Image and Video: Part-II – Designs and Applications. *IEEE Trans. on Image Proc.*, 12:696–705, 2003.

[26] YouTube. http://www.youtube.com/.

[27] P. R. Zimmermann. *The Official PGP User's Guide*. The MIT Press, 1995.